



UNIVERSIDADE DA CORUÑA



Model-based methods for small area estimation. Habilitation Thesis. Tomas Hobza

July 3, 2017

This Habilitation Thesis summarizes main results of the candidate (sometimes we will also refer as “author”) obtained in Small Area Estimation (SAE). This memory can be divided into two parts. In the first part, the basic concepts are introduced and the reader is placed in the concrete models that are going to be treated, indicating the developments obtained in the current literature and his contributions. In the second part, the author presents the contributions in the form of published papers.

In this report the candidate shows a very knowledgeable about the existing literature in this area, and clearly indicates the main ideas. It is remarkable the wide knowledge of the computational problems that the fitted of these models entails. Providing solutions of great interest in the study of real problems, particularly in the field of official statistics or socio-economic studies. The rigorousness of the presented studies is emphasized, in which not only theoretical results of relevance are obtained, but they are supported by complete simulation studies, to finish showing the application of the methodology developed in a real case. Finally, in the section of conclusions and future lines, we can see the author's projection, with proposals of study that pose new challenges, expanding and improving the proposed estimators in the current literature.

In particular, and following the division that the author makes, we will comment on each of the parts of his work.

Models for continuous responses

These models are very studied in the literature. And, as the author says, they can be divided into two types of models depending on the information that you have. The author presents 5 main contributions: 3 articles in JCR journals and 2 chapters of the same book.

Main strengths: the contribution of methodology for both models, individual models and area models. The author studies the case where the domains are divided into two groups and have different behaviour within them. One more step that the author gives is to consider that the area effects can be fixed or random for some areas. Closing the contributions of this part with the small area estimation under random regression coefficient models, only for the case of individual models. Highlighting the applications on databases of socio-economic interest, such as the study of proportion of unemployed people by sex in the Canary Islands and the estimation of poverty proportions in the

Spanish provinces (using the Living Conditions Survey). These works show the difficulty of the fitted algorithms and the estimators, and the importance to work under the correct model.

There are no significant weaknesses. As a minor remark, although the candidate publishes in JCR journals relevant to the subject, it would be advisable to publish in JCR journals of greater impact.

Questions and comments:

- In these papers the importance of working under the correct model is shown, it would be interesting to consider some test to study the type of effects in the model (fixed or random) and then apply the methodology developed in these works.
- The R-language computational implementation of this methodology is of great interest to the scientific community. Do you plan to put in a repository, or available to those interested in the topic?

Models for discrete responses

These models are of great interest in situations where observations are discrete or categorical. For example, if a person is unemployed or not, if you are at risk of poverty or not, if you are a smoker or not,... Many of these real problems are sometimes modeled as continuous variables (the previous case). Nevertheless, the quantities of interest are usually proportions or counts, therefore models for discrete responses are more appropriate. The main contributions of the author in this line of work are 5 papers in JCR journals and 1 paper in review.

Main strengths: The contributions are for the unit level model, studying the empirical best predictor (EBP) for the model with random effect of area and its extension to more complex models that consider the random effect of area and time. The obtaining of these estimators implies a great methodological ability in SAE and fitting algorithms. Its computational implementation is also complex, because the usual packages do not make these calculations and it becomes necessary to implement fitting algorithms that work well. Again, the author presents complete simulation studies that validate the theoretical results and the application to a problem of social interest, estimation of poverty proportions in the counties of the region of Valencia in Spain.

In this section, but in another line of work, robust estimation and outlier detection is presented with important applications to SAE. Consistency and asymptotic normality of the median estimators and in other types are proved in his works. These works, with important theoretical results, are complemented by simulation studies.

There are no significant weaknesses.

Questions and comments:

- From the results it follows that the EBP and Plug-in have similar behavior. But the theoretical properties and the difficulties to implement in practice are very different. So, based on the author's experience, which one would you recommend for use in practice?
- It would be interesting to see the gain of using these models versus the models of continuous response, from the previous chapter, in the examples of poverty or unemployment study that were studied.

- The R-language computational implementation of this methodology is of great interest to the scientific community. Do you plan to put in a repository, or available to those interested in the topic?

For all the aforementioned, Thomas Hobza is a candidate with a very good research trajectory and with a good future projection. As a reviewer of this work, I believe that the candidate is highly qualified for the position of assistant professor.

He has developed relevant works of great interest in the field of statistics, with important applications in official statistic, as evidenced by the publications in relevant journals, research reports and reviewed articles. Furthermore, the candidate presents a great future projection, contributing to scientific growth and training potential researchers, which are the guarantee of future in the university.

Sincerely,

**LOMBARDIA
CORTIÑA MARIA
JOSE -
33322690Z**

Firmado digitalmente por LOMBARDIA
CORTIÑA MARIA JOSE - 33322690Z
Nombre de reconocimiento (DN): c=ES,
serialNumber=33322690Z,
sn=LOMBARDIA CORTIÑA,
givenName=MARIA JOSE,
cn=LOMBARDIA CORTIÑA MARIA JOSE -
33322690Z
Fecha: 2017.07.04 12:51:32 +02'00'

María José Lombardía Cortiña
Department of Mathematics